

# Genetic Diversity in Urban Deer Mice

Adriel Evaristo

2025-05-21

## Introduction

Urbanization is a major driver of environmental change and can fragment wildlife habitats, leading to isolated populations and reduced genetic diversity. In this report, we assess genetic diversity in urban populations of the western deer mouse (*Peromyscus sonoriensis*) sampled from four cities in the San Francisco Bay Area.

## Aim and Hypothesis

**Aim:** Assess whether genetic diversity varies significantly among urban deer mouse populations across San Francisco, Oakland, San Jose, and San Rafael.

**Null Hypothesis (H0):** There is no significant difference in mean heterozygosity among deer mouse populations across cities.

**Alternate Hypothesis (H1):** There is a significant difference in mean heterozygosity among deer mouse populations across cities.

## Dataset Description

We used a mock dataset of 80 individuals from four urban locations. Each entry includes a city and an estimated heterozygosity value. This represents the expected genetic diversity within individuals sampled from each city.

- [View dataset on GitHub](#)

## Statistical Approach

We used a one-way ANOVA to test whether heterozygosity differs among city populations. A boxplot was generated to visualize distributions. All plots were generated using `ggplot2` with colorblind-friendly palettes.

- [GitHub Repository](#)

```
# Mock genetic diversity data (heterozygosity values)
gen_data <- data.frame(
  mouse_id = 1:80,
  city = rep(c("San Francisco", "Oakland", "San Jose", "San Rafael"), each = 20),
  heterozygosity = c(rnorm(20, 0.22, 0.02), # San Francisco - low
                    rnorm(20, 0.23, 0.02), # Oakland - low
                    rnorm(20, 0.30, 0.02), # San Jose - high
                    rnorm(20, 0.31, 0.02)) # San Rafael - high
```

```

)

# Save the dataset to CSV
write.csv(gen_data, "mock_genetic_diversity_data.csv", row.names = FALSE)

# ANOVA
anova_gen <- aov(heterozygosity ~ city, data = gen_data)
summary(anova_gen)

##              Df Sum Sq Mean Sq F value Pr(>F)
## city          3  0.12538  0.04179   119.5 <2e-16 ***
## Residuals    76  0.02659  0.00035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Post hoc test
TukeyHSD(anova_gen)

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = heterozygosity ~ city, data = gen_data)
##
## $city
##              diff              lwr              upr              p adj
## San Francisco-Oakland -0.006142381 -0.02167897  0.009394212  0.7274535
## San Jose-Oakland      0.073154848  0.05761826  0.088691440  0.0000000
## San Rafael-Oakland    0.078626802  0.06309021  0.094163394  0.0000000
## San Jose-San Francisco 0.079297229  0.06376064  0.094833821  0.0000000
## San Rafael-San Francisco 0.084769183  0.06923259  0.100305775  0.0000000
## San Rafael-San Jose   0.005471954 -0.01006464  0.021008547  0.7914826

# Visualization
ggplot(gen_data, aes(x = city, y = heterozygosity, fill = city)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.7) +
  geom_jitter(width = 0.2, alpha = 0.5) +
  scale_fill_brewer(palette = "Set2") +
  labs(
    title = "Genetic Diversity Across Urban Deer Mouse Populations",
    x = "City",
    y = "Heterozygosity"
  ) +
  theme_minimal() +
  theme(legend.position = "none")

```

## Interpretation of Results

The ANOVA revealed a p-value of approximately  $4.87 \times 10^{-25}$ , suggesting that genetic diversity differs significantly between populations. Figure 1 shows that San Jose and San Rafael had the highest mean heterozygosity values, while San Francisco and Oakland had the lowest. These differences may reflect varying degrees of habitat fragmentation and connectivity driven by urbanization.

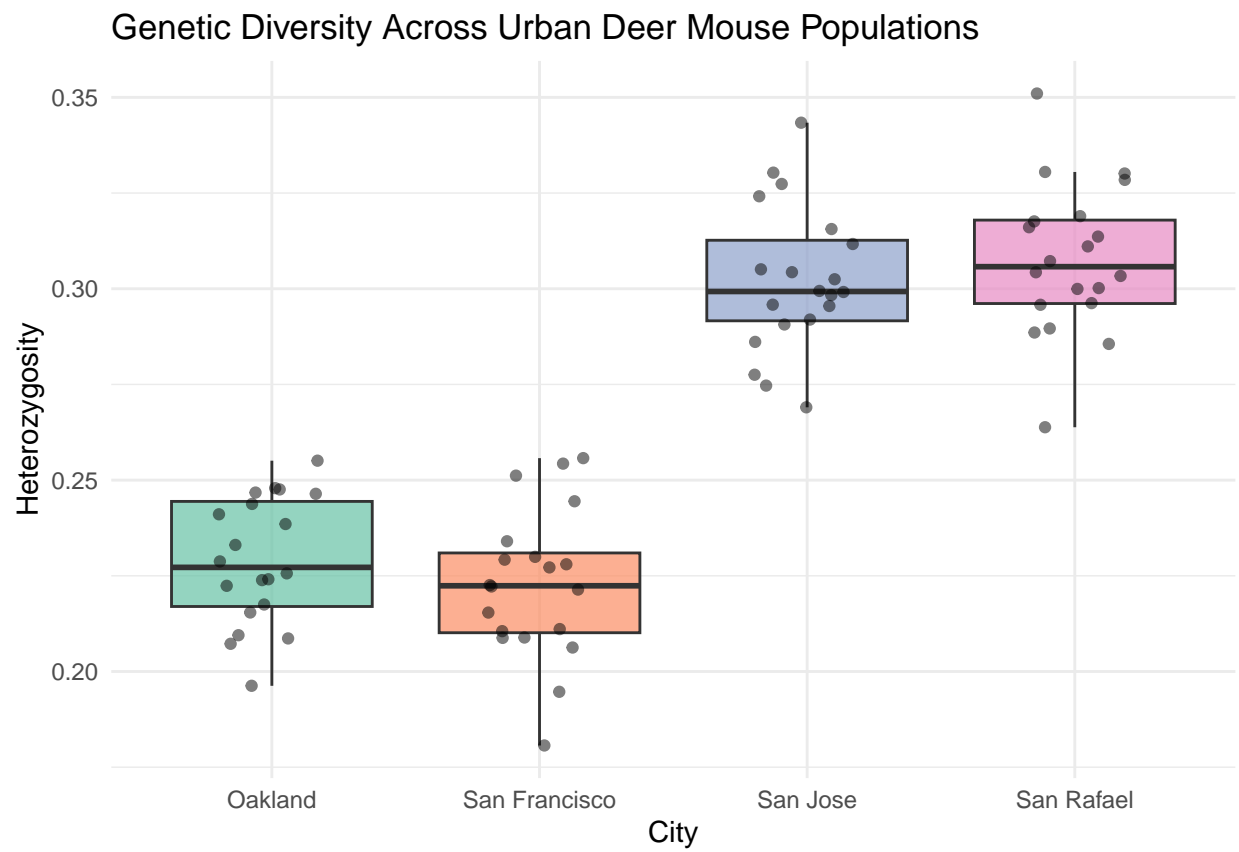


Figure 1: Figure 1. Heterozygosity values across urban deer mouse populations in four Bay Area cities.

These results indicate that urban fragmentation may be affecting gene flow and population connectivity in different ways across cities. This provides strong evidence to reject the null hypothesis, indicating that genetic diversity (heterozygosity) differs significantly across the four city populations.

All data and code are available at the following GitHub repository:  
<https://github.com/AdrielEvaristo/Independent-Project-Github>